Using ISI's 'Highly Cited Researchers' to obtain a Country Level Indicator of Citation Excellence

Aparna Basu

aparnabasu.dr@gmail.com
Institute of Genomics and Integrative Biology, *located at TCGA,* 254 Okhla Industrial Estate, New Delhi 110020, (India)

Abstract
A high level of citation to an author's work is, in general, a testimony to the fact that the author's work has been noted and used by his peers. High citation has been found to be correlated with other forms of recogtion and rewards, and is a key indicator of research performance. The Institute for Scientific information (ISI) defines a 'highly cited researcher' (*HCR*) as one of 250 most cited authors of journal papers in any discipline. Citation data for 20 years (1981-99) is used to calculate the share of *HCR*s for countries in 21 subject areas. We find that the US dominates in all subject areas (US share ~ 40-90%). Based on the number of highly cited researchers in a country, an index of citation excellence is proposed. We find that rank order of countries based on this index is in conformity with our general understanding of research excellence, whereas the more frequently used indicator, citations per paper, gave an unacceptable rank order due to an inherent bias toward very small countries. Additionally, a high value of the index of citation excellence was found to be associated with higher concentration of highly cited researchers in affiliating organizations.

Abbreviations
CP = Citations per paper
HCR = Highly Cited Researcher
HCRP = HCRs per paper
IHCR = Intensity of HCRs in affiliating institutions
IoCE = Index of Citation Excellence

Introduction

Recently, ISI-Thompson Scientific has made available data on the most highly cited researchers, based on its analysis of 19 million papers and 5 million author names in the Science Citation Index for the 20 year period 1981-99. 'Highly cited researchers' (hereafter referred to as *HCR*) are defined by ISI as being among the 250 most cited researchers in a particular subject area. For 21 broad subject areas as defined by ISI, this list when complete will contain the names of more than top 5000 authors. (*A citation is a reference made to work of an author in another scientific article as an acknowledgement of prior work.*) While *HCRs* constitute a mere 0.1 percent of all publishing scientists ('scientists' is used here in a generic sense and includes social scientists), it is an important segment that contributes to the leading edge of science. This makes it a very valuable source of information on persons whose work has been frequently noted by peers (ISIHighlyCited.com, 2004).

While citations to journal papers are a measurable and useful indicator of the use to which a piece of research has been put, (despite the variability in citation patterns across disciplines and with time), there is a debate on how well citation frequencies correlate in general with other forms of recognition such as peer assessment (Warner (2000)). This is due to problems such as language, self-citation, eponymous references to work without corresponding citation, delayed recognition etc. (Garfield (1980), Haggbloom (2002) , Glanzel et al (2003, 2004),). However, there are studies showing that the citation levels of *top cited scientists* correlate well with accepted forms of recognition (Tijssen (2002)) and rewards such as prestigious awards, memberships of scientific academies etc. Garfield has studied the correlation between high citation frequency and the receipt of prestigious prizes, especially the Nobel Prize. (Garfield (1992, 1984)). ISI's 'highly cited researchers' database, therefore, not only provides an opportunity for identifying individuals with high performance and potential, it has created a corpus of information that can form a baseline for scientometric and sociological studies on research excellence for many years to come.

Batty (2003) had obtained the country-wise distribution of highly cited researchers using ISI data of 2002. However, his data consisted of only 100 cited individuals in 14 research fields (it

Proceedings of ISSI 2005, pages 352–361.
Edited by P. Ingwersen and B. Larsen.

excluded mathematics, social sciences and humanities). Since then the ISI data on highly cited researchers is being continuously expanded to cover 250 individuals in 21 fields (*in April 2004 it covered more than 4500 author names*). Batty had also shown that there was strong geographical localization of institutions to which highly cited researchers were affiliated. They were mainly concentrated on the East and West coasts of America, and in the UK, Europe and Japan, and followed a power law distribution.

The main objective of our study is to extend Batty's work by finding the share of highly cited researchers in different countries for the expanded ISI list in all 21 fields. Assuming that highly cited researchers represent "excellence" in science, we have used the country-share to define an '*index of citation excellence*' (*IoCE*) and compared it to another citation based indicator at the country level. We have also obtained the 'intensity' of highly cited researchers in affiliating organizations (*IHCR*). We test for a relationship, if any, of concentration in affiliating organizations, *IHCR,* and the index of citation excellence *IoCE*.

Data and methodology

Data on *HCR*s in 21 different fields (disciplines) was downloaded from the ISI database on April 15, 2004. The data was analyzed to obtain the country distribution of *HCR's* in each discipline. (*We have used a fractional count for researchers who are highly cited in more than one field*.) The numbers for each discipline were then aggregated to obtain the country distribution of *HCR*s in all fields combined.

The relationship between the share of highly cited researchers in a country and its productivity (publication output) has been explored using regression.

The number of *HCR*s in each country was then adjusted for size of country output by normalizing by the publication output *P* (productivity). We obtained the number of "highly cited researchers per paper" (*HCRP*) for a country,

$$HCRP\ (Country\ X) = \frac{Numberofhighlycitedresearchers HCR(CountryX)}{Numberofpublishedpapers P(CountryX)} \qquad (1)$$

Data on scientific productivity is based on Science and Engineering Indicators (2002).

The "index of citation excellence" *IoCE,* has been defined here as,

$$IoCE\ (Country\ X) = \frac{ShareofHCRs(CountryX)}{Shareofpublications(CountryX)} - 1 \qquad (2)$$

$$= \left\{ \frac{HCRsofCountryX \,/\, TotalHCRs}{PublicationsofCountryX \,/\, Worldpublications} \right\} - 1 \qquad (3)$$

The World average for the index of excellence is zero. Countries with positive values for *IoCE* have more than the world average of *HCR*s adjusted for size of country output.

Citations per paper, *CP* is

$$CP\ (Country\ X) = \frac{Totalcitations(CountryX)}{Totalpublications(CountryX)} \qquad (4)$$

Country rankings for citations per paper are based on Essential Science Indicators (2002).

The number of *HCR*s per affiliating institution (a*n affiliating institution is defined here as one that has at least one highly cited researcher,*) gives the "intensity of HCR's" (*IHCR*) as,

$$IHCR\ (country\ X) = \frac{Numberofhighlycitedresearchers HCR(CountryX)}{Numberofaffiliatinginstitutes(CountryX)} \qquad (5)$$

Affiliating institutions are assumed to be representative of the reputed centres of research in a country. All data have been processed using Microsoft Excel.

Distribution of Highly Cited Researchers in Countries worldwide

Based on ISI data of April 2004, we find that there are 4569 highly cited researchers distributed in 587 institutions in 38 countries (*39 including Korea, which could not be located on the ISI site*). The USA dominates world science with 3082 HCR's in 587 institutions. Other countries with more than 100 HCR's are the UK (354 in 106 institutions), Germany (194 in 113 institutions), Japan (176 in 81 institutions), Canada (144 in 35 institutions), and France (117 in 64 institutions) (see Table 1). (*The earlier study by Batty covered 1222 scientists working in 429 institutions located in 27 countries*). Countries with more than 100 highly cited '*affiliating institutes'* (i.e., those institutes with at least one highly cited researcher) are the US with 587, Germany with 113 and the UK with 106 institutes.

Table 1: Distribution of Highly Cited Researchers and Affiliating Institutes by Country (1989-1991)

(1)	(2)	(3)	(4)	(5)
Country	**Cited Institutions**	**Highly cited Researchers (HCRs)**	**Intensity HCR/Inst**	**Rank by HCRs**
US	587	3082	5.25	1
UK	106	354	3.34	2
Germany	113	194	1.72	3
Japan	81	176	2.17	4
Canada	35	144	4.11	5
France	64	117	1.83	6
Australia	26	78	3	7
Switzerland	29	75	2.59	8
Netherlands	30	64	2.13	9
Italy	32	46	1.44	10
Sweden	13	44	3.38	11
Israel	9	37	4.11	12
Belgium	12	24	2	13
Denmark	12	23	1.92	14
New Zealand	5	15	3	15
Spain	9	12	1.33	16
Austria	9	10	1.11	17
PR China	6	9	1.5	18
India	7	8	1.14	19
Finland	5	7	1.4	20
Norway	3	7	2.33	20
S. Africa	3	7	2.33	20
Russia	2	5	2.5	23
Taiwan	3	5	1.67	23
Hungary	2	4	2	25
Singapore	1	4	4	25
Greece	2	3	1.5	26
Mexico	3	3	1	26
Brazil	2	2	1	28
Chile	2	2	1	28
Poland	1	2	2	28
Algeria	1	1	1	29
Iran	1	1	1	29
Pakistan	1	1	1	29
Panama	1	1	1	29
Phillipines	1	1	1	29
Portugal	1	1	1	29
Total	**1222**	**4569**	**3.74**	

We note that there is considerable variation in intensity between countries. This feature will be dealt with in more detail in a later section.

Relationship between Highly Cited Researchers and Productivity

We note that the distribution of highly cited researchers is highly skewed, with a few countries accounting for the HCRs. How does a country's share of HCR's compare with its share of publication output? Here we explore the relationship between HCRs (y) and the total scientific productivity of countries(x). Figure 1 shows the best fit; a quadratic relation ship with a high level of significance ($y = 0.0523x2 + 0.247x + 0.0835$; $R^2 = 0.9282$)

The relationship implies that the probability for a country to have highly cited researchers scales as the second power of the country's total production of papers.

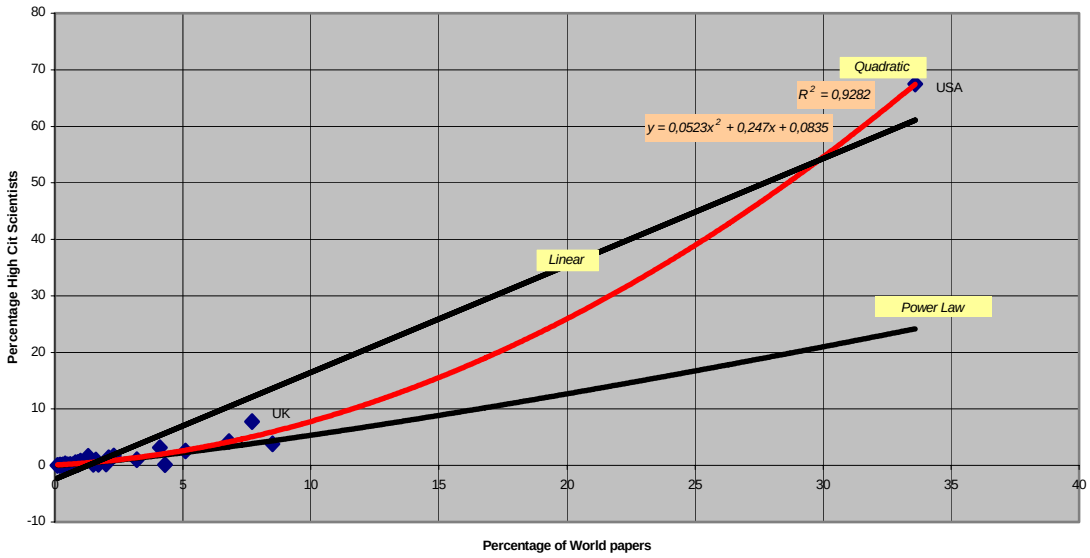


Figure 1. Highly cited researchers in a country scale as the square of total scientific productivity of the country ($R^2 = 0.93$)

Distribution of Highly Cited Researchers by Country and Discipline

ISI defines 'highly cited researchers as the top 250 cited researchers in 21 different disciplines or broad subject fields. The subject fields used by ISI are listed in Table 2.

Table 2: List of Disciplines used by ISI

Agricultural Sciences	Mathematics
Biology & BioChemistry	Microbiology
Chemistry	Molecular Biology & Genetics
Clinical medicine	Neuroscience
Computer Science	Pharmacology
Ecology/Environment	Physics
Economics/Business	Plant & Animal Science
Engineering Geosciences	Psychology/ Psychiatry Social Sciences
Immunology	Space Sciences
Materials Science	

The ranked list of countries based on the number of HCRs in each discipline, is shown in Table 3. Russia and Brazil are not included as they are not among the top 10 countries in any discipline. The US, UK and Canada are in the top ten list of countries in all 21 fields, and Germany, Japan France and the Netherlands in more than 15 fields. Israel and Sweden, ranked lower than 10 overall, are ranked third in Computer Science and Social Sciences respectively.

In terms of share, the US has the by far the highest percentage of HCR's in every subject area. Only in 3 subject areas (namely Pharmacology, Agriculture and Plant and Animal Sciences) does the US have less than half of all highly cited scientists. The average US figure for all subjects combined is about 67%. In Social Sciences, Business/Economics, Psychology/Psychiatry and Clinical Medicine, the proportion of HCR's living and working in the US is higher than 90percent. Apart from these last four subjects, the UK has approximately 5-10% of highly cited researchers in all other subjects. Germany has more than 10% HCR's in Chemistry and Japan in Agriculture and Biology and Biochemistry. (Figure 2)

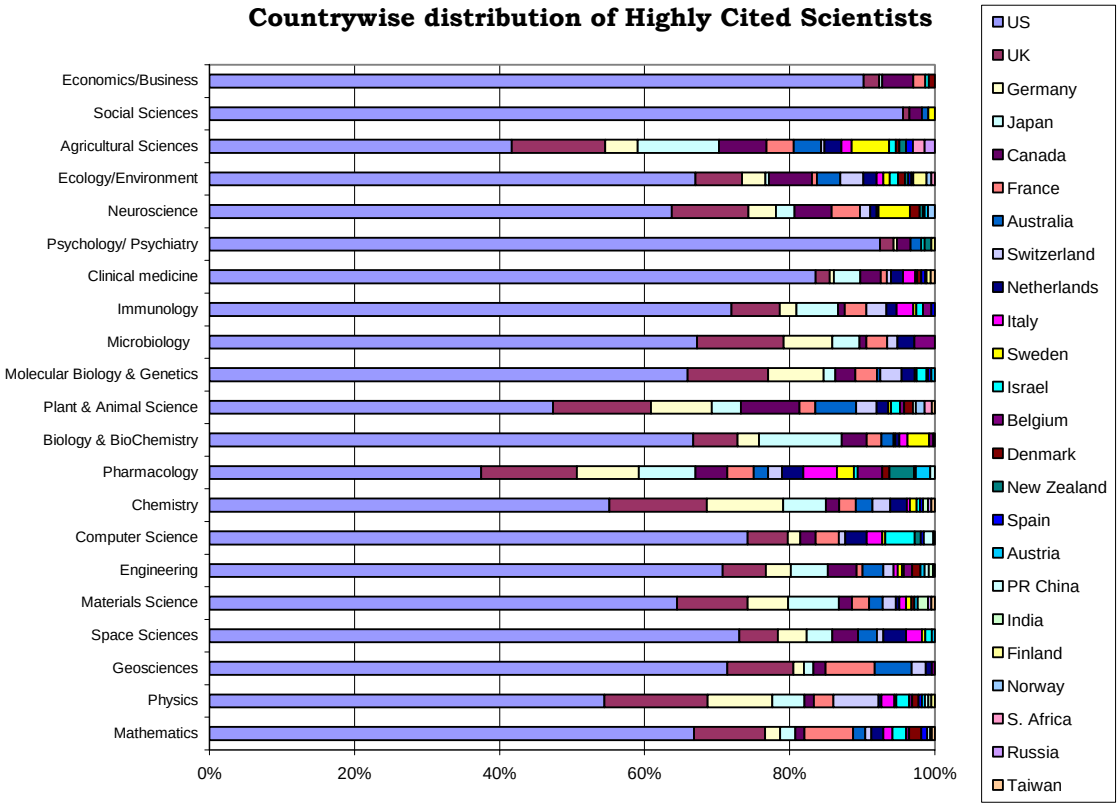
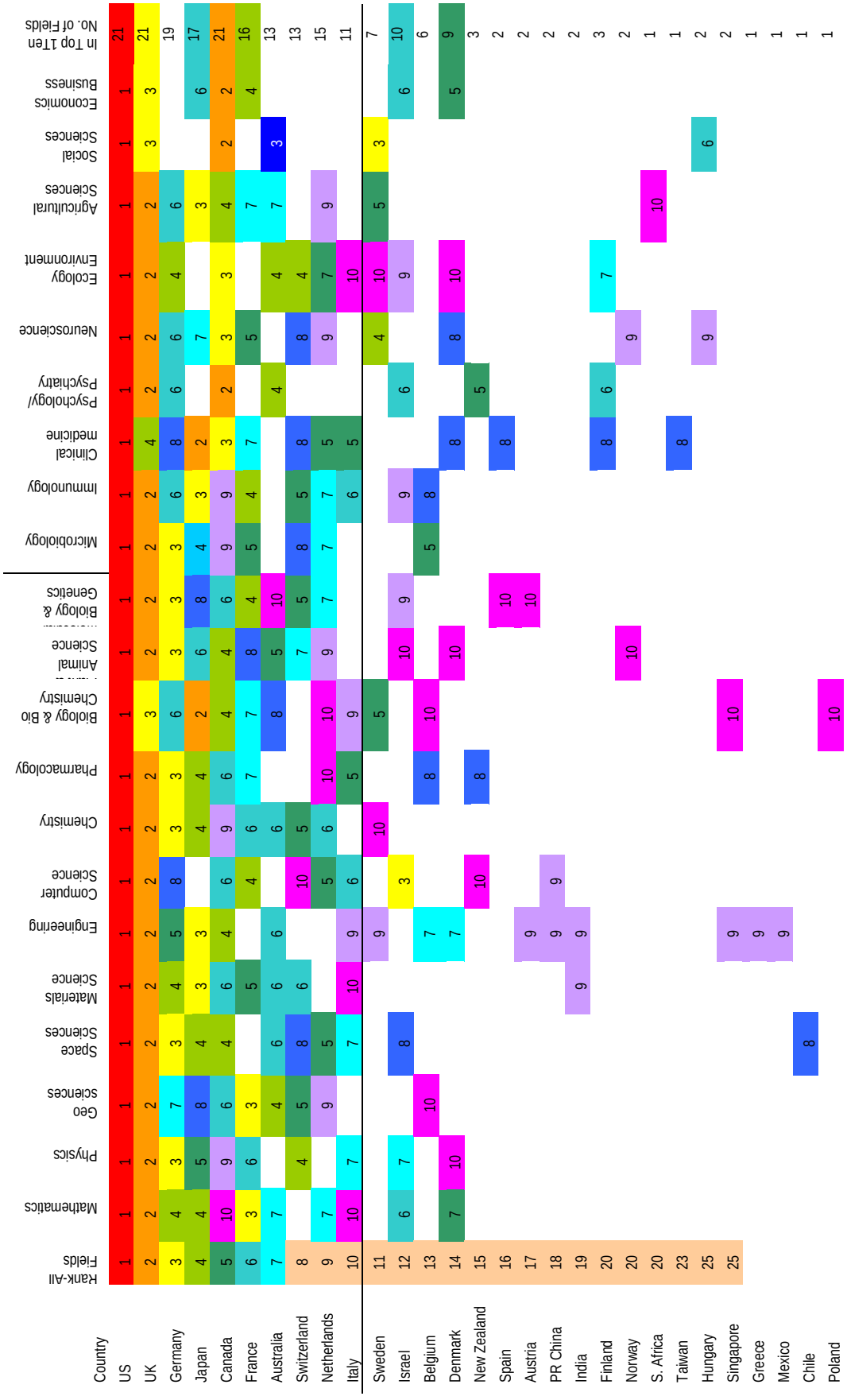


Figure 2. The distribution of Highly Cited Researchers in different disciplines by country.

Table 3: Country Rank for Number of Highly Cited researchers (by Discipline)



Indicator of Citation Excellence - IoCE

The trend of increasing HCRs with output of a country (Fig.1) needs to be removed before making comparisons between countries. To adjust for the substantial size differences between countries, we normalize the number of HCRs of a country by its total scientific productivity per annum. As defined earlier (Eqn.2), the "*Indicator of Citation Excellence*" (**IoCE**) is then,

$$IoCE\ (Country\ X) = \frac{\text{Percentage of "highly cited researchers" of X}}{\text{Percentage of World papers of X}} - 1$$

with positive values indicating an index better than world average (world *IoCE*=0). The country share and ranking based on HCRs and IoCE are shown in Table 2.

Table 4: Country Share of Highly Cited Researchers and Rank

(1)	(2)	(3)	(4)	(5)	(6)
Country	**% of World HCRs**	**Rank by HCR**	**IoCE**	**Rank by IoCE**	**Rank by total citations(*) Top 20**
USA	67.45	1	*1.008*	*1*	1
Switzerland	1.64	8	*0.263*	*2*	9
UK	7.75	2	*0.006*	*3*	2
New Zealand	0.33	15	*-0.179*	*4*	-
Israel	0.81	12	*-0.190*	*5*	16
Canada	3.15	5	*-0.231*	*6*	6
Australia	1.71	7	*-0.258*	*7*	10
Netherlands	1.40	9	*-0.333*	*8*	8
Denmark	0.50	14	*-0.371*	*9*	17
Germany	4.25	3	*-0.376*	*10*	3
Sweden	0.96	11	*-0.398*	*11*	11
Belgium	0.53	13	*-0.416*	*12*	14
France	2.56	6	*-0.498*	*13*	5
Japan	3.85	4	*-0.547*	*14*	4
Singapore	0.09	26	*-0.562*	*15*	-
S. Africa	0.15	20	*-0.617*	*16*	-
Austria	0.22	17	*-0.635*	*17*	-
Italy	1.01	10	*-0.685*	*18*	7
Norway	0.15	20	*-0.694*	*19*	-
Hungary	0.09	25	*-0.708*	*20*	-
Chile	0.04	29	*-0.781*	*21*	-
Finland	0.15	20	*-0.781*	*21*	18
Iran	0.02	32	*-0.781*	*21*	-
Mexico	0.07	27	*-0.781*	*21*	-
Greece	0.07	27	*-0.836*	*25*	-
PR China	0.20	18	*-0.869*	*26*	19
Spain	0.26	16	*-0.869*	*26*	12
Taiwan	0.11	23	*-0.878*	*28*	-
Portugal	0.02	32	*-0.891*	*29*	-
India	0.18	19	*-0.897*	*30*	-
Brazil	0.04	29	*-0.937*	*31*	-
Poland	0.04	29	*-0.945*	*32*	-
Russia	0.11	23	*-0.975*	*33*	15

HCR= highly cited researchers, IoCE= Indicator of Citation Excellence; Period 1981-1999; (*) Source: Essential Science Indicators (2002)

The world average for the indicator *IoCE* being set at zero, we find three countries having values of *IoCE* better than the world average, viz. the USA, Switzerland and the UK.

The only country which features among the top 15 countries by total number of HCR's, but not in the top 15 countries by *IoCE* is Italy (ranked 10 by *HCR* and 18 by *IoCE*). Singapore is ranked at 15 with 0.2 percent of the world's production of papers and 0.1 percent of *HCR*s.

It may be noted that the ranking based on citation (column 6, Table 4) has a fair correlation with the ranking based on highly cited researchers (column 5, Table 1). The ranking based on the index *IoCE* (Column 5, Table 4), however, highlights countries whose share of *HCR*s is higher than expected from their productivity levels, such as Switzerland, New Zealand, Israel and Denmark.

It is possible that language plays a role in the high citation levels of authors with a bias toward English speaking countries such as Canada, Australia, New Zealand.

Index of Citation Excellence as compared to Citations per Paper

Here we show that the Index of Citation Excellence *IoCE*, while accounting for size differences between countries, is a more reliable indicator of performance as compared to *CP*, the number of citations per paper which is also normalized by the size of a country's output.

In Table 4, we show a list of the top 15 countries ranked by citations per paper (Essential Science Indicators, 2003)

Table 4: Countries ranked by Citations per Paper (*)

Rank	COUNTRY	PAPERS	CITATIONS	CITATIONS PER PAPER
1	GUINEA BISSAU	116	2,070	17.84
2	BERMUDA	185	2,711	14.65
3	RWANDA	179	2,517	14.06
4	SENEGAMBIA	383	5,373	14.03
5	SWITZERLAND	129,785	1,585,691	12.22
6	USA	2,618,154	30,765,049	11.75
7	PANAMA	900	10,031	11.15
8	CONGO DEM REP	412	4,308	10.46
9	ICELAND	2,800	29,137	10.41
10	NETHERLANDS	184,526	1,908,540	10.34
11	CONGO PEOPL REP	72	734	10.19
12	DENMARK	72,630	735,002	10.12
13	SWEDEN	144,425	1,446,651	10.02
14	GAMBIA	485	4,832	9.96
15	ENGLAND	570,667	5,628,105	9.86
16	SCOTLAND	88,836	873,438	9.83
17	CANADA	346,126	3,259,935	9.42
18	FINLAND	66,524	610,841	9.18
19	BELGIUM	92,266	825,768	8.95
20	SEYCHELLES	88	772	8.77

* Source: http://in-cites.com/countries/2002allfields.html

It is clear from Table 4 that 'citations per paper' (CP) is not a good indicator of national excellence because of the undue weight received by small countries with very low productivity (less than 1000 papers). This is a problem frequently encountered with ratio based indicators. However, since there are only 33 countries with highly cited scientists *(as opposed to over 100 countries with citations*), and none of them have very low levels of production, this problem is not encountered when computing the ratio of 'highly cited researchers' to 'papers'. Therefore, we claim that the index *IoCE*, or '***index of citation excellence***' based on the number of highly cited researchers per paper, is a more suitable indicator of excellence at the country level than the more commonly used indicator 'citations per paper'.

Concentration of HCR's in affiliating institutes versus Excellence

Batty's work on citation geography highlights the heavy concentration of institutions to which HCRs were affiliated in certain locations around the globe, chiefly in cities on the East and West coasts of the USA, and in UK, Europe and Japan. The affiliating institutions are the centres of excellence as commonly understood. We have found a similar concentration of highly cited researchers in some institutions, thus taking Batty's study of geographical concentration to the next level, namely institutional concentration. While Batty's 2003 study only considered 1222 scientists, our study considers almost four times that number.

In this section we have explored the relationship, if any, of the indicator of excellence *IoCE* (y) with concentration (intensity) of highly cited scientists in affiliating institutions (x) using regression. The country data and fitted linear equation are seen in Figure 3. ($y = 0.29x - 1.17$; $R^2=0.59$), the relation explaining about 60% of the variance in the data.

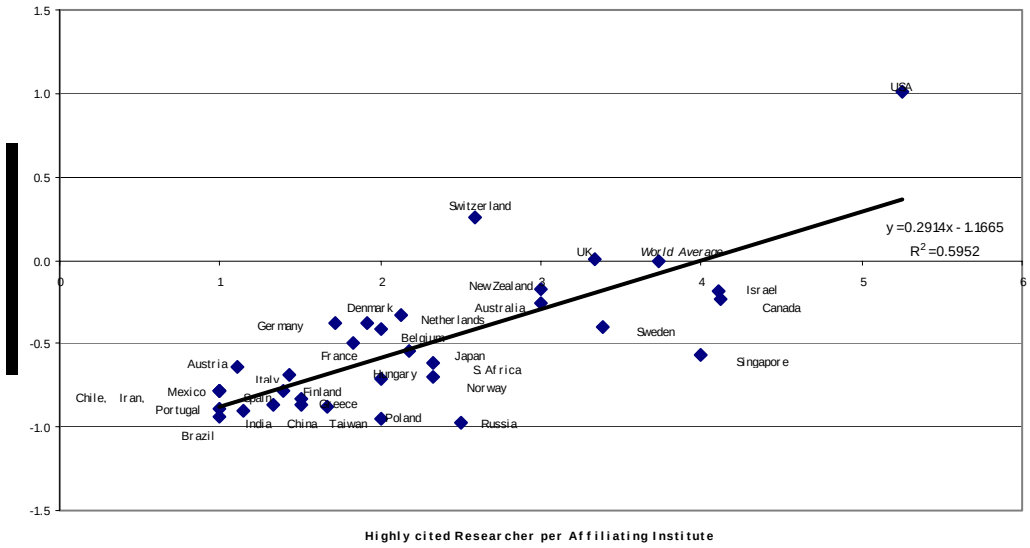


Figure 3. Positive correlation between the Indicator of Citation Excellence and the intensity of Highly Cited Researchers per affiliating institute (R^2= 0.595)

Two significant outliers are the USA and Switzerland with *IoCE* much higher than that predicted by the fit. Another outlier is Russia with the number of cited researchers normalized by total output being much lower than that predicted by the relationship.

Summary and Conclusions

Our results are based on ISI Thompson Scientific data on 'highly cited researchers' (HCRs) in the 20 year period 1981-1999. Comparability with inferences derived from Essential Science Indicators (ESI) which covers 10 year data periods may be problematic. We have found that the **USA** houses two thirds of the world's highly cited researchers (67%), while producing only one third of the world scientific literature (33%) and receiving about half of the world's citations. This is another example of the

Matthew effect (Merton, 1968). The next three countries ranked by share of highly cited researchers are the **UK** (7.8%), **Germany** (4.3%), and **Japan** (3.9%). The USA has more highly cited researchers than all other countries put together in all subjects but three (**Pharmacology, Agriculture, Plant and Animal Sciences**, where the US share is less than 50%). In **Social Sciences, Business /Economics, Psychology/Psychiatry and Clinical Medicine**, the US share is higher than 90%.

We found that the probability that a country will have highly cited researchers increases in proportion to the square of its own research output. To account for variation in scientific output of countries, we have normalized the share of *HCR*s by the country's share of publications. This puts the **USA**, **Switzerland**, **UK** in the top positions. We have shown that a new indicator proposed by us, the **Index of Citation Excellence (IoCE)** is a more reliable indicator of research performance at the country level than '**Citations per paper'**, as the latter yields unreasonably high ranks for small countries with very low productivity, and is therefore unreliable as an indicator of country performance. This problem is avoided in the case of IoCE, as very small countries do not have highly cited researchers and drop out of the ranking altogether.

We have also found a positive correlation between institutional concentration of *HCR*s and excellence (using *IoCE* as an indicator.) This may be a reflection of selection effects or an indication that centres of excellence attract together individuals with high levels of performance.

While highly cited researchers appear to provide a good indicator of research performance of a country, there are certain other considerations that need to be taken into account if one is to make valid international comparisons. The mobility of researchers who often come to a country with high level qualifications from their country of origin is one such consideration. How does one apportion research credit under these circumstances? Should the country of residence get the full credit for the researcher? A significant numbers of persons of foreign origin are found in the highly cited researchers' lists in some developed countries (Batty 2003, Basu, 2004). Some of these issues will have to be addressed with if we are to define excellence in terms of a manpower based quantifiable parameter.

References

Basu, A. (2004) Country Indicators of Excellence based on ISI's "highly cited researchers", *Poster, 8th International S&T Indicators Conference*, Leiden, Sep. 23-24, 2004 (Powerpoint)

Batty**,** M (2003) Citation Geography: It's about location, The Scientist, 17(16), Aug. 25

Batty, M (2003) The geography of scientific citation, Environment Planning A, 35: 761-5

Essential Science Indicators (2002) http://in-cites.com/countries/2002allfields.html

Garfield, E. (1992) A review of several of these studies appears on his website: http://www.garfield.library.upenn.edu/essays/v15p116y1992-93.pdf

Garfield, E. (1984) The Awards of Science: Beyond the Nobel Prize. Part 2.The Winners and Their Most-Cited Papers, Essays of an Information Scientist, Vol:7, p.405-419, *Current Contents*, #50, p.3-17, December 10, 1984

Garfield, E. (1980) Premature discovery or delayed recognition--Why?" *Curr Contents*, 21:5-10

Glanzel W. and E. Garfield (2004) The Myth of delayed recognition, *The Scientist*, 18 (11) 8, 2004

Glänzel, W. et al., (2003) "Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon," *Scientometrics*, 58:571-86

Haggbloom, Steven J. (2002) The 100 Most Eminent Psychologists of the 20th Century, *Review of General Psychology*, Vol. 6, No. 2, 139–152

ISI 'highly cited researchers', www.isihighlycited.com, data collected April 15, 2004

Merton, Robert K (1968), The Matthew Effect in Science, Science, 159(3810): 56:63, Jan 5

NSF Science and Engineering Indicators (2000) Vol. 2, National Science Board, USA. More recent data available at http://www.nsf.gov/sbe/srs/seind04/start.htm

Tijssen, Robert J. W., Martijn S. Visser, Thed N. van Leeuwen (2002) Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics,* **54** (3): 381-397

Warner, J. (2000) A critical review of the application of citation studies to the Research Assessment Exercises, Journal of Information Science, 26 (6) 2000, pp. 453–460